A Scalable Solution Partially Supervised Approach for Generation of Family Signature against Android Malware

J. Ramya and K. Chandramohan

Abstract--- Clustering has been well studied for desktop malware analysis as an effective triage method. Malware could be malicious software that gets installed in your device and performs unwanted tasks. Mainly designed to transmit information about your web browsing habits to the third party Conventional similarity-based clustering techniques, however, cannot be immediately applied to Android malware analysis due to the excessive use of third-party libraries in Android application development and the widespread use of repackaging in malware development. Deceitful practices in Google Play, the most prominent Android application showcase, fuel look rank mishandle and malware multiplication. To distinguish malware, past work has concentrated on application executable and authorization examination. The proposed algorithm for a user should be given reviews to the app at one time again the user cannot be given to the review. After downloading mobile applications from Google play users are asked to give reviews about that particular downloaded applications. However fraudulent developers give fake ratings, about their application promote their application to the top. Hoax web algorithm [HWA] identified that for the detection of the ranking, rank, and review based evidence are considered.

Keywords--- Third-Party, Clustering Techniques, Fraud Application, Hoax web Algorithm.

I. INTRODUCTION

Malware could be as software that performing actions intended by an attacker without the consent of the owner when executed. Each malware has a specific characteristic, attack goal, and propagation method. Five main categories of malware types are a virus, worm, Trojan horses, backdoors and spyware. The digital mobile telecommunication is now in its third decade and is steadily progressing. The advancement in telecom has touched all its components including the mobile stations. There are different malware detection techniques such as signature-based malware detection, specification-based detection, anomaly-based detection, and machine learning based detection. Malware detection is a significant topic over the World Wide Web. Malware or Malicious Software could be as software designed to distort and interrupt the mobile or computer applications, collect relevant information and hence perform malicious operations. Operations include over private information, covertly steal this over the system, display annoying advertisement, the users.

This can be extended to other potential financial transactions that are made through SMS as well. As for the Android-specific vulnerabilities, the problem consists of two design features relating to the way SMS messages are sent and received on Android. A new Android malware detection approach is developed using parallel machine learning classifiers.

Several security firms across the globe are busy preparing patches and cures for the plethora of malware existent nowadays. However, the fact remains that for every

J. Ramya, Student, Computer Science and Engineering, Gnanamani College of Technology, Pachal, Namakkal.

K. Chandramohan, M.E, Ph.D., Assistant Professor, Computer Science and Engineering, Gnanamani College of Technology, Pachal, Namakkal.

cure created for malware, a subtle variant of the same malware is created that bypasses all the latest security patches thereby reversing all the hard work and effort put in to counter them. To make things worse, malware is becoming smarter every day, and polymorphic malware are the latest entrants in this calamitous game of defeating the opponent.

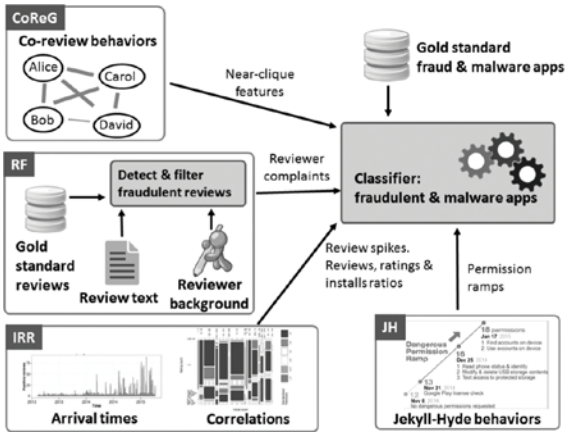

Figure 1.1: Malware Detection

Malware Types

To have a better understanding of the technique and logic behind the malware, to use to classify it. Malware will be divided into several classes depending on its purpose. The levels are as follows:

Virus

This is the simplest form of software. It is just any piece of software that is loaded and launched without user's permission while reproducing itself or infecting (modifying) other software.

Worm

This malware type is very alike to the virus. The difference is that a worm will Spread over the network and replicate to other machines.

Trojan

This malware class is used to define the malware types that aim to appear as legitimate software. Because of this, the general spreading vector utilized in this class is social engineering, i.e., making people think that they are downloading legitimate software.

Adware

The only purpose of this malware type is displaying on the computer. Frequently adware will be as seen as a subclass of spyware, and it will very unlikely lead to dramatic consequence.

Spyware

As it implies from the name, the malware that performs espionage will be referred to as spyware. Typical actions of spyware include tracking search history to send personalized advertisements, tracking activities to sell them to the third parties consequently.

Rootkit

Its functionality permit the attacker to access the data with higher authority than is allowed. For example, it will be used to give an unauthorized user administrative access. Rootkits always hidden its existence and quite often are unnoticeable on the system, making the detection and therefore removal incredibly hard.

Backdoor

The backdoor could be a type of malware that provides an extra secret "entrance" to the system for attackers. It does not cause any harm but issue attackers with a broader attack surface. However of this, backdoors are never used independently. They are preceding malware attacks of another type.

II. RELATED WORKS

Android malware is increasingly extended in terms of complexity. To evade signature-based detection, that represents the most adopted method by present antimalware vendors, malware writers begin to deploy malware with the ability to change their code as they propagate [2]. They are generalized and are not user specific. For such type of problems, Rank Fusion of Results from Multiple Search Engines is the solution [5]. Available Meta Search Engines though access a large number of conventional search

engines, they do not consider a specific user's preferences [1]. Cloud data owners prefer to outsource documents in an encrypted form for privacy preserving. At the same time, a valid mechanism is also proposed to guarantee the correctness and completeness of search results. Besides, we analyze the search efficiency and security under two popular threat models [10].The static features are obtained by analyzing an executable file. Deep learning-based malware detection has so far focused on analyzing executable files and runtime API calls [12]. Static code analysis approaches face challenges due to obfuscated code and adversarial perturbations [7].Generally, malware is defined as a security threat for computing systems and computer networks. Today's common and most used method for detecting malware is a signature-based method. There is also another way that focuses on file behavior[11]. Evaluation results indicate low resilience of the antimalware detection engines against code obfuscation [4]. Furthermore, we evaluate the resilience of Androguard's code similarity and AndroSimilar's robust statistical feature signature against code obfuscated malware. It presented Dalvik byte code obfuscator to evaluate the resilience of anti-malware techniques against Android byte code transformations prevalent on x86. Evaluation results suggest that the effectiveness of commercial anti-malware should be further investigated[5].In the similarity identification phase, the location and page similarity is identified by computing similarity among the locations and retrieval pages [3]. In the computation of frequent access pattern, find all the frequent-retrieval of the web pages by computing the support value. The Web Search engines provide results based on the given query by the user [8]. On the other hand, for a similar query, there is a different context because every user has different interests. Due to the enormous growth of the information on the web, retrieving accurate information for satisfying customers is a complicated task for the web search engines[14]. However, malware attackers increasingly employ techniques such as repackaging and obfuscation to bypass signatures and defeat attempts to analyze their inner mechanisms. The increasing sophistication of Android malware calls for new defensive techniques that are harder to evade, and are capable of protecting users against new threats [11].Mobile devices do not have the resources of desktops and cannot support a full anti-malware system. Created a system that inspects and detects mobile-network traffic. The solution sits at a network gateway, similar to an IDS for a traditional network. To reduce the amount of time to process the network traffic and not slow it, the significant reduction is used. Web search personalization using collaborative filtering adapts a generic search engine for the needs of a community of users. It makes use of click-through data and concept clusters obtained from web snippets which are captured at the backend. Collaborative filtering is employed for ranking. Despite the numerous efforts to thwart the growth of mobile malware, the number of mobile malware is getting increased by evolving themselves. By applying, for example, code obfuscation or junk code insertion, mobile malware can manipulate its appearance while maintains the same functionality. Thus mobile malware can easily evade the existing anti-mobile-malware solutions. Malware Behavior Feature (MBF) extraction method, and proposes the malicious behavior feature based malware detection algorithm. Finally, we designed and implemented the MBF based malware detection system, and the experimental results show that it can detect newly appeared unknown malware. The content-based ranking is based on contents and keywords rather than link structure and keywords provided by search engines. Search engines results are retrieved based on the user query. Also, the usage-based ranking algorithm considers the past user navigation pattern and analyze the behavior of the user to recommend the data. The personalization provides content and services based on the knowledge about an individual's preferences and behavior. Malware has been effectively classified based on their families. Using the classified malware system and detected the malware that is found during the routing of files between clients and server. In this

work a malware classification system has been developed using approximate matching of data flow graphs with string signatures. Using an aggregated signature versus many signatures for each variant for detection of malware provides many benefits to anti-virus vendors and the community as a whole by reducing the size of the signature database, reducing maintenance, and increasing the speed of detection without losing accuracy. To protect and secure information from malware attacks, many different techniques are being identified and proposed by researchers. Malware detection and classification is a challenging area of research as a large number of new malware variants are introduced day by day. Metamorphic malware causes another challenge as it varies structurally with every new infection. The commonly used signature-based malware detection fails in detecting mobile malware most of the times. The studies reveal that behavioral or heuristic approach is more effective for detection of metamorphic malware. The malicious executable is analyzed using two approaches: disassembly, utilizing IDA-Pro, and the application of a dedicated state machine to obtain the set of functions comprising the executable. The signature extraction process is based on a comparison with a common function repository. By eliminating functions appearing in the common function repository from the signature candidate list, F-Sign can minimize the risk of false-positive detection errors.This signature is support on the malware's execution data extracted from kernel of objects, and neither uses malicious code syntax specific information code execution flow information. The signature is more resistant to obfuscation methods and resilient in detecting malicious code variants. Search engines have greatly influenced the way people access data on the Internet as such engines give the preferred entry point to billions of pages on the Web. However, highly ranked web pages generally have higher visibility to people and pushing the ranking higher has become the top priority for webmasters. PageRank, which enables the classified tree to be constructed according to a large number of users' similar searching results, and can reduce the problem of Theme-Drift, caused by using PageRank only, and the problem of outdated web pages. It improves the searching efficiency without reducing the searching speed, which provides the users with sufficient expanded information relevant to searching content. The collaborative algorithm to optimize the user profile parameter vector was also proposed to improve efficiency. The re-ranking of results driven by the user's interests in various web page features improves the responsiveness of the search engine to the preference and intent of the user, which is the primary motive of semantic search.

III. METHODS AND IMPLEMENTATION

Due to the increasing openness and popularity, Android phones had been an attraction to most of the malicious applications and an attacker will easily embed its own code into the code of a benign application. Therefore, malware attacking the android application an alarming rate and under these circumstances, security of the devices and allow access to, be at stake. Static detection methodologies based on signature-based approaches that are widely used in the Android platform to detect malicious applications. Accurately detect malware by extracting signatures from test information and then comparing the test benign samples. Traditional methods like signature-based ones cannot protect users from the ever-increasing sophistication and rapid behavior changes of new types of Android malware. However, lots of recent efforts have been made to use machine learning to characterize and discover the malicious behavior patterns of mobile apps for malware detection.

The personality feature of mobile devices, malware detection, is critical and is a must tool in each device to protect private data and attack. In this paper proposed a system to analyze different malware detection used for search rank fraud mobile operating system. The documentation uploaded by the creator is seeable to the user. The fraud application is identified using rating analysis, and through this, we tend to return to understand

whether the request is

Ornot. Developers give fake ratings and promote they are the top. Hoax web algorithm [HWA] identified that for the detection of the ranking, rating, and review based evidence are considered.

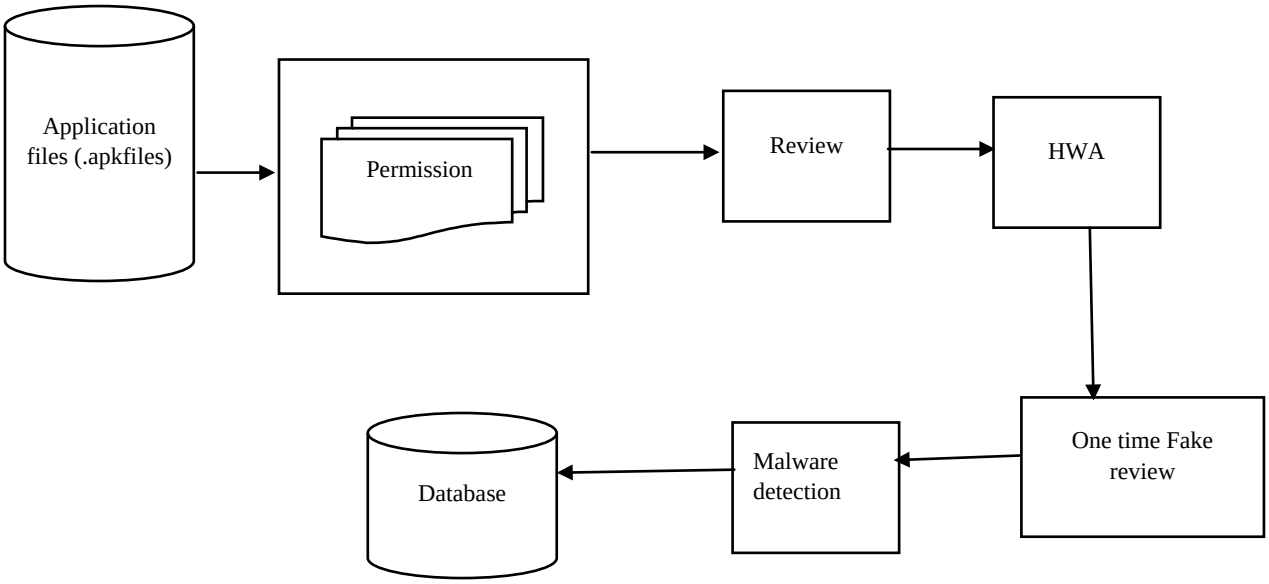

Figure 3.1: Implementation Proposed Architecture

Text-based Malware Detection in Fake Review

Signatures is a proven technology, but exponential growth in unique malware programs has caused an explosion in signature database sizes. It is important to provide security solutions to these devices before these threats cause widespread damage. The considerations for malware detection on devices and propose a signature-based malware detection method. Specifically, we detail a proposed algorithm that is well suited for use when a fake review is sent to the user by review and asks the user for personal information through mobile or web link.

Signature-based Detection

Signature-based methods primarily focus on categorizing a code as malware based on the signatures by comparing it with existing malware family signatures. It is highly efficient but fails to detect the unknown malware due to the limited size of the signature database. Although hash signatures had allow false positive rate, the number of malware samples covered by each hash signature is also small – typically one. As an outcome, the total size of the hash signature set grows with the exponential growth in the number of unique malware samples.

Malware Detection in Review

Malware is a computer program which is designed to harm or disrupts the operation of a computer system. Malware detection is a system that attempts to find whether an application has malicious intent or not. The review in malware detection to send the details or link, for example, you've won the lottery or a contest you didn't enter, then unfortunately for you, it's probably a hacking attempt.

Hoax web algorithm

After downloading mobile applications from Google play users are asked to give that particular downloaded applications.

Developers give fake ratings, promote they're the top. Hoax web algorithm [HWA] identified that for the detection of the rank ranking, evaluation, and review based evidence are considered.

Hoax algorithms steps are given to below:

Step 1: download the app

Step 2: a review of the app

Step3: users are assigned to fake ratings the developer identifies the fake ratings.

Step 4: put review = (review);

Getreview=user review TimeStamp

Set daydiff=(now – DateHours ("yyyy-mm-dd", h));

```
if review <= 1 {
        msg= msg .show ("successfully upload review ")
        return
    }
elseifreview >=1
{
        msg=msg. show ("fake review")
        return
    }
    }
```

IV. Result And Discussion

Malware could be malicious software that gets installed in your device and performs unwanted tasks. Mainly designed to transmit data about your web browsing habits to the third party Conventional similarity-based clustering method, But, cannot be immediately applied to Android malware analysis due to the excessive use of third-party libraries in Android application development and the widespread use of repackaging in malware development using asp.net software tool to development the detection of the system and improve the page rank. Fair Play correlates review activities and uniquely combines detected review relations with linguistic and behavioral signals.

Average Page Rank and Detection of Malware

To measure the relative importance of web pages. There as on that Page Rank is interesting that there is any case where simple counting.

There is a small problem with this ranking function. Consider two web pages that point to each on another button another page.

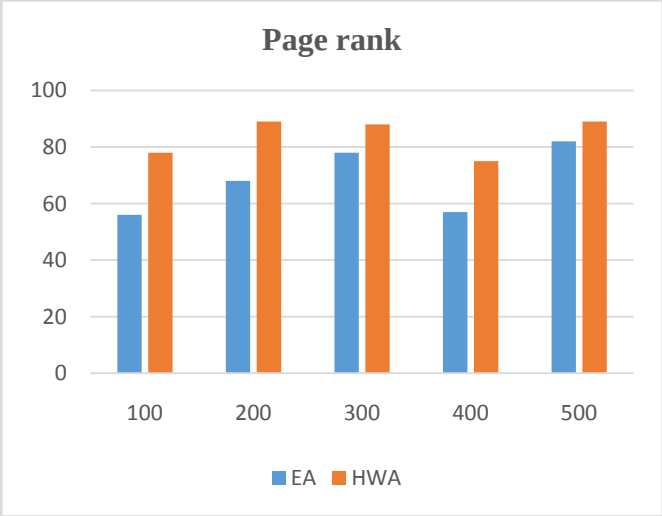

Figure 4.1: Comparison of Page Rank

Number of Apps Reviewed

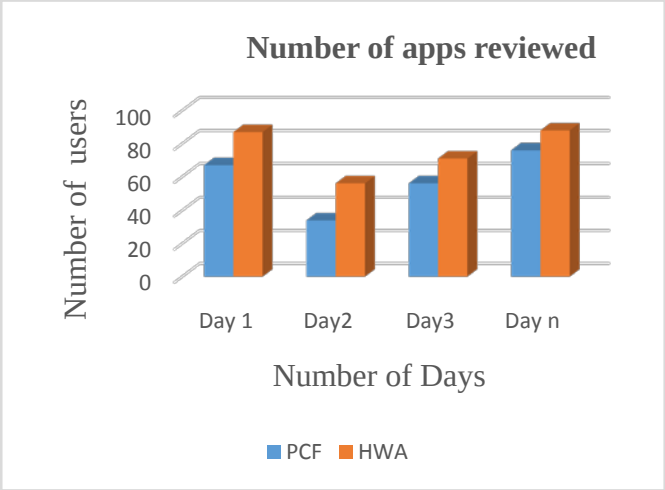

Figure 4.2: Number of app reviewed

The apps to stores are more competitive every day, and app publishers are looking for more intelligence and organic advantages to build a sustainable mobile business. This means investing in tools like Appetitive for in-app feedback and retention and poring over in-app analytics, attempting to glean essential insights that will unlock another step up in growth.

Detect Fraudulent Development Fake Ratings

Many users reviewed by the number of apps. Review timestamps should have a one-time granularity. Suppose the users gave review more than one time it is denoted by fake rating.

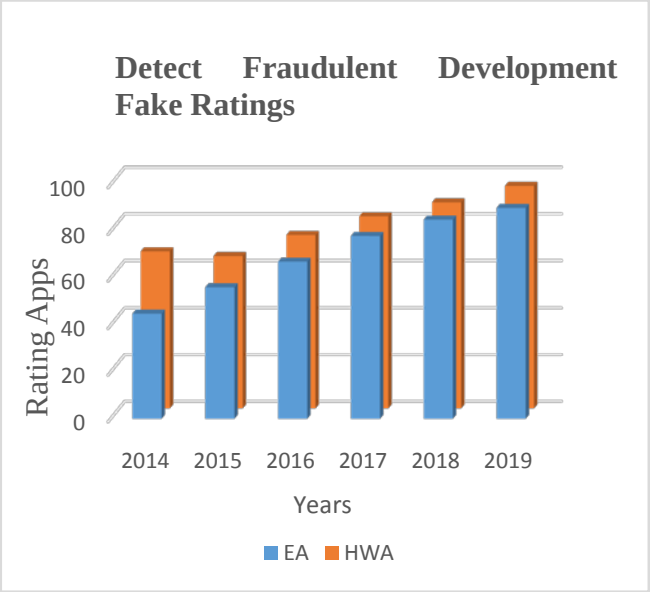

Figure 4.3: Comparisons of Rating

V. CONCLUSION

Malware could be malicious software that gets installed in your device and performs unwanted tasks. Mainly designed to transmit the data about your web browsing habits to the third partyConventional similarity-based clustering techniques. To distinguish malware, past work has concentrated on application executable and authorization examination. Hoax web algorithm [HWA] identified that for the detection of the rank ranking, rating, and review based evidence are considered. The proposed algorithm for a user should be given reviews to the app at one time again the user cannot be given to the review. After Downloading mobile applications from Google play users are asked to give reviews about that particular downloaded applications.

REFERENCES

[1] A. Rahul and R. Soni, "Rank Fusion of Results from Multiple Search Engines: An Implementation", International Conference on Machine Intelligence and Research Advancement, Pp. 224-229, 2013.

[2] C. Gerardo, A. Di Sorbo, F. Mercaldo and C. Aaron Visaggio, "Obfuscation techniques against signature-based detection: a case study", Mobile Systems Technologies Workshop (MST), Pp. 21-26, 2015.

[3] C. Chi, X. Zhu, P. Shen, J. Hu, S. Guo, Z. Tari and A.Y. Zomaya, "An efficient privacy-preserving ranked keyword search method", IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 4, Pp. 951-963, 2016.

[4] C. Li, S. Sultana and R. Sahita, "Henet: A deep learning approach on intel® processor trace for effective exploit detection", IEEE Security and Privacy Workshops (SPW), Pp. 109-115, 2018.

[5] D. Hamid, B. Pechaz and M. Vafaie Jahan, "Malware detection using markov blanket based on opcode sequences", International Congress on Technology, Communication and Knowledge (ICTCK), Pp. 564-569, 2015.

[6] F. Parvez, V. Laxmi, V. Ganmoor, M. Singh Gaur and A. Bharmal, "Droidolytics: robust feature signature for repackaged android apps on official and third party android markets", 2nd International Conference on Advanced Computing, Networking and Security, Pp. 247-252, 2013.

[7] F. Parvez, A. Bharmal, V. Laxmi, M. Singh Gaur, M. Conti and M. Rajarajan, "Evaluation of android anti-malware techniques against dalvik bytecode obfuscation", IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, Pp. 414-421, 2014.

[8] F. Parvez, A. Bharmal, V. Laxmi, V. Ganmoor, M. Singh Gaur, M. Conti and M. Rajarajan, "Android security: a survey of issues, malware penetration, and defenses", IEEE communications surveys & tutorials, Vol. 17, No. 2, Pp. 998-1022, 2015.

[9] F. Shunkai, J. Zhang and R. Mu, "Ranking factors in devising practical POI search model", Proceedings IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Pp. 267-272, 2011.

[10] H. Shifu, A. Saas, L. Chen and Y. Ye, "Deep4maldroid: A deep learning framework for android malware detection based on linux kernel system call graphs", IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), Pp. 104-111, 2016.

[11] H. Neminath and H. Dogra, "Detecting Packed Executable File: Supervised or Anomaly Detection Method?", 11th International Conference on Availability, Reliability and Security (ARES), Pp. 638-643, 2016.

[12] H. Kelly and Y. Qu, "Performance Measures of Behavior-Based Signatures: An Anti-malware Solution for Platforms with Limited Computing Resource", Ninth International Conference on Availability, Reliability and Security, Pp. 303-309, 2014.

[13] D. Indumathi and A. Chitra. A collaborative search with query expansion and result re-ranking", World Congress on Information and Communication Technologies, Pp. 985-989, 2011.

[14] D. Kamini, M. Suresh and S. Neduncheliyan, "Encrypted multi-keyword ranked search supporting gram based search technique", International Conference on Information Communication and Embedded Systems (ICICES), Pp. 1-6, 2016.

[15] K. Jonghoon, J. Jeong, J. Lee and H. Lee, "Droidgraph: discovering android malware by analyzing semantic behavior", IEEE Conference on Communications and Network Security, Pp. 498-499, 2014.

[16] L. Wu, P. Ren, K. Liu and H. Duan, "Behavior-based malware analysis and detection", First International Workshop on Complexity and Data Mining, Pp. 39-42, 2011.

[17] L. Yan, Y. Li, M. Xu and W. Hu, "A user model based ranking method of query results of meta-search engines", International Conference on Network and Information Systems for Computers, Pp. 426-430, 2015.

[18] M. Foram, J.R. Aishwarya, V. Panchal and V. Pinjarkar, "Hybrid crawling for time-based personalized web search ranking", International conference of Electronics, Communication and Aerospace Technology (ICECA), Vol. 1, Pp. 252-255, 2017.

[19] S. Muthurajkumar, M. Vijayalakshmi, S. Ganapathy and A. Kannan, "Agent based intelligent approach for the malware detection for infected cloud data storage files", Seventh International Conference on Advanced Computing (ICoAC), Pp. 233-237, 2015.

[20] Q. Yanzhen and K.Hughes, "Detecting metamorphic malware by using behavior-based aggregated signature", In World Congress on Internet Security (WorldCIS-2013), Pp. 13-18, 2013.

[21] R. Mehul Smriti, "Analysis of desktop search and ranking of their results based on semantics from user feedback",12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Pp. 241-245, 2016.

[22] B. Haniye Razeghi and M. Abadi, "MalHunter: Automatic generation of multiple behavioral signatures for polymorphic malware detection", In ICCKE, Pp. 430-436, 2013.

[23] K. Sasidharan Satheesh and C. Thomas, "A Survey on Metamorphic Malware Detection based on Hidden Markov Model", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Pp. 357-362, 2018.

[24] S. Asaf, E. Menahem and Y. Elovici, "F-sign: Automatic, function-based signature generation for malware", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 41, No. 4, Pp. 494-508, 2011.

[25] S. Raja Khurram and N. Lavesson, "Veto-based malware detection", Seventh International Conference on Availability, Reliability and Security, Pp. 48-54, 2012.

[26] F. Shosha, Ahmed, C. Ching Liu, P. Gladyshev and M. Matten ,"Evasion-resistant malware signature based on profiling kernel data structure objects", 7th International Conference on Risks and Security of Internet and Systems (CRiSIS), pp. 763-771, 2012.

[27] S. Vedpal, A. Chaudhary and P. Punia, "OSA-PR: Optimized searching algorithm based on page ranking: Proposed algorithm", IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA), pp. 233-243, 2012.

[28] S. Mingshen, X. Li, J.C.S. Lui, R. TB Ma and Z. Liang, "Monet: a user-oriented behavior-based malware variants detection system for android", IEEE Transactions on Information Forensics and Security, Vol. 12, No. 5, Pp. 1103-1112, 2017.

[29] C. Tian, "A kind of algorithm for page ranking based on classified tree in search engine", International Conference on Computer Application and System Modeling (ICCASM 2010), Vol. 13, Pp. V13-538, 2010.

[30] V. Ganesh, and A. Ravichandran, "Adaptive Semantic Search: Re-Ranking of search results based on Webpage feature extraction and implicitly learned knowledge of User Interests", 10th International Conference on Semantics, Knowledge and Grids, Pp. 345-351, 2014.